



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### **Community-led comparative genomic and phenotypic analysis of the aquaculture pathogen *Pseudomonas baetica* a390T sequenced by Ion semiconductor and Nanopore technologies**

**Citation for published version:**

Beaton, A, Lood, C, Cunningham-Oakes, E, MacFadyen, A, Mullins, AJ, Bestawy, WE, Botelho, J, Chevalier, S, Coleman, S, Dalzell, C, Dolan, SK, Faccenda, A, Ghequire, MGK, Higgins, S, Kutschera, A, Murray, J, Redway, M, Salih, T, da Silva, AC, Smith, BA, Smits, N, Thomson, R, Woodcock, S, Welch, M, Cornelis, P, Lavigne, R, van Noort, V & Tucker, NP 2018, 'Community-led comparative genomic and phenotypic analysis of the aquaculture pathogen *Pseudomonas baetica* a390T sequenced by Ion semiconductor and Nanopore technologies', *FEMS Microbiology Letters*.  
<https://doi.org/10.1093/femsle/fny069>

**Digital Object Identifier (DOI):**

[10.1093/femsle/fny069](https://doi.org/10.1093/femsle/fny069)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

FEMS Microbiology Letters

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



**Ainsley Beaton<sup>1\*</sup>, Cédric Lood<sup>2,3\*</sup>, Edward Cunningham-Oakes<sup>4\*</sup>, Alison MacFadyen<sup>10\*</sup>, Alex J. Mullins<sup>4\*</sup>, Walid El Bestawy<sup>1</sup>, João Botelho<sup>13</sup>, Sylvie Chevalier<sup>11</sup>, Shannon Coleman<sup>7</sup>, Chloe Dalzell<sup>1</sup>, Stephen K. Dolan<sup>7</sup>, Alberto Faccenda<sup>1</sup>, Maarten G. K. Ghequire<sup>2</sup>, Steven Higgins<sup>8</sup>, Alexander Kutschera<sup>14</sup>, Jordan Murray<sup>1</sup>, Martha Redway<sup>1</sup>, Talal Salih<sup>1</sup>, Ana C. da Silva<sup>9</sup>, Brian A. Smith<sup>5</sup>, Nathan Smits<sup>3</sup>, Ryan Thomson<sup>1</sup>, Stuart Woodcock<sup>12</sup>, Martin Welch<sup>6</sup>, Pierre Cornelis<sup>11</sup>, Rob Lavigne<sup>3</sup>, Vera van Noort<sup>2</sup>, Nicholas P. Tucker<sup>1</sup>**

1. Strathclyde Institute of Pharmacy and Biomedical Science, University of Strathclyde, 161 Cathedral Street, Glasgow, G4 0RE, United Kingdom

2. Centre of Microbial and Plant Genetics, KU Leuven, Belgium

3. Laboratory of Gene Technology, KU Leuven, Belgium

4. Cardiff School of Biosciences, Cardiff University, Sir Martin Evans Building, Park Place, Cardiff, CF10 3AX, United Kingdom

5. The University of Arizona, School of Plant Sciences, P.O. Box 210036, Forbes Building, 303 Tucson, Arizona, 85721-0036, United States

6. Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom

7. Lower Mall Research Station, 2259 Lower Mall, University of British Columbia, V6T 1Z4, Vancouver, BC, Canada

8. Department of Plant and Microbial Biology, University of Zürich, Switzerland

9. Centre for Biomolecular Sciences, School of Life Sciences, University of Nottingham, Nottingham, United Kingdom

10. Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush Campus, Midlothian, Scotland, United Kingdom

11. Laboratoire Microbiologie Signaux et Microenvironnement (LMSM), Université de Rouen, France

12. John Innes Centre, Norwich, United Kingdom

13. UCIBIO/REQUIMTE, Laboratório de Microbiologia, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal

14. Phytopathology, Center of Life and Food Sciences, Weihenstephan, Technical University of Munich, Germany

*\*These authors contributed equally to the sequencing or to intensive whole genome data analysis of P. baetica. AB generated semiconductor short reads whilst CL produced Nanopore long reads. ECO and AM performed the phylogenetic analysis whilst AMac performed whole genome BLAST analysis and editing.*

## Abstract.

*Pseudomonas baetica* strain a390T is the type strain of this recently described species and here we present its high-contiguity draft genome. To celebrate the 16<sup>th</sup> International Conference on *Pseudomonas*, the genome of *P. baetica* strain a390T was sequenced using a unique combination of Ion Torrent semiconductor and Oxford Nanopore methods as part of a collaborative community-led project. The use of high quality Ion Torrent sequences with long Nanopore reads gave rapid, high contiguity and quality, 16 contig genome sequence. Whole genome phylogenetic analysis places *P. baetica* within the *Pseudomonas koreensis* clade of the *Pseudomonas fluorescens* group. Comparison of the main genomic features of *P. baetica* with a variety of other *Pseudomonas* spp. suggests that it is a highly adaptable organism, typical of the genus. This strain was originally isolated from the liver of a diseased wedge sole fish and genotypic and phenotypic analyses show that it is tolerant to osmotic stress and to oxytetracycline.

## Introduction.

In September 2017, the biannual conference on the biology of *Pseudomonas* was held in Liverpool, UK. All aspects of *Pseudomonas* biology were reported at the meeting from clinical to environmental microbiology. The vast majority of the *Pseudomonas* literature focuses on just a few organisms, most notably *Pseudomonas aeruginosa*, *Pseudomonas fluorescens*, *Pseudomonas putida* and *Pseudomonas syringae* although the genus is far more diverse than this. In order to contribute to addressing this imbalance, the genome of *Pseudomonas baetica*, a recently described member of the genus was sequenced for the *Pseudomonas* 2017 genomics forum. The resulting genome sequence has been analysed by members of the *Pseudomonas* 2017 community from across the globe using Slack as a collaboration tool (Perkel 2016).

*Pseudomonas baetica* strain a390 is one of five Gram-negative organisms that were isolated from a fish disease outbreak in an aquaculture facility in Huelva, Spain in 2006 (López *et al.* 2012). The infected fish were wedge sole (*Dicologlossa cuneata*), a flat fish that is widely consumed in Andalucía and parts of France where there is increasing interest in sustainable aquaculture of this species. Aquaculture not only provides a sustainable source of protein but it is also an economically important industry around the world. European commission data reveals that 1.25 million tonnes of food are produced by aquaculture in the EU every year and that over 85,000 people are directly employed in the industry (European Commission 2014). A disadvantage to aquaculture systems is that the unusually high

population density of the farmed fish leads to an elevated risk of infectious disease outbreaks. For this reason, antibiotics such as oxytetracycline are routinely used as growth promoters in aquaculture fish food pellets (Leal *et al.* 2017). Other members of the *Pseudomonadaceae* are well known to cause disease in aquaculture scenarios, most notably *Pseudomonas anguilliseptica* is a pathogen of both farmed and wild eels (Joh *et al.* 2013).

*P. baetica* was formally described in 2012 using a combination of 16S rRNA sequencing and phenotypic analyses, allowing a biochemical profile to be generated (López *et al.* 2012). This analysis revealed that *P. baetica* is a member of the *P. fluorescens* group and is particularly closely related to *P. koreensis*. Although infected wedge sole showed no obvious visual signs of *P. baetica* infection, intraperitoneal injection and immersion infection assays yielded 100% and 10% mortality rates, respectively (López *et al.* 2012). Subsequent infection assays with a variety of other fish species demonstrated that *P. baetica* caused higher mortality in wedge sole than other species including sea bass and sea bream and that there is a temperature-dependent effect on virulence (López *et al.* 2017). In order to rapidly identify *P. baetica* contamination in an aquaculture setting, Lopez *et al.* have developed immunological and molecular detection assays (López *et al.* 2017).

Although *P. baetica* was originally described from an aquaculture infection scenario, a number of recent studies have identified this species and close relatives such as the newly described *Pseudomonas helmanticensis* and *Pseudomonas granadensis* in diverse environments in the rhizosphere including bean roots in Iran, and forest soils in Spain (Ramírez-Bahena *et al.* 2014; Pascual *et al.* 2015; Keshavarz-Tohid *et al.* 2017). This suggests that *P. baetica* is more likely to be an opportunistic pathogen of fish as opposed to a genuine marine pathogen. We hypothesised that the draft genome sequence of *P. baetica* would reveal the genetic basis for the traits required to cause disease in wedge sole. The aim of this study is to provide genomic insights into the biology of *P. baetica* and to use this genome as a reflection on the diverse interests of the *Pseudomonas* community.

## Materials and Methods

### Bacterial strains and genome sequence accessions

*Pseudomonas baetica* a390T originally isolated from the liver of wedge sole fish, *Dicologlossa cuneate*, and described by Lopez *et al.* (2012), was obtained from DSMZ (DSM no. 26532) and is considered to be the type strain. Unless stated otherwise, all phenotypic comparisons were performed against *Pseudomonas fluorescens* PF01, *Pseudomonas putida* KT2440 and *Pseudomonas aeruginosa* PA14. This Whole Genome Shotgun project

has been deposited at DDBJ/ENA/GenBank under the accession PKLC000000000. The version described in this paper is version PKLC010000000. Raw sequencing reads were also deposited and can be found with accession numbers SRR6792524 and SRR6792523 for the Nanopore and Ion Torrent reads respectively.

## **DNA extraction, library preparation and Ion Torrent PGM Sequencing**

DNA extraction was carried out using ISOLATE II genomic DNA kit (BIOLINE) using the standard protocol for bacterial cells. Fragmentation was carried out using the standard protocol from NEBNext Fast DNA Fragmentation & Library Prep Set for Ion Torrent and adapter-ligated DNA was prepared using barcode set 6. Size selection was carried out using the E-gel ibase (Invitrogen), and analysed using the 2100 Bioanalyzer (Agilent Genomics). A fragment size of 424bp was carried forward. Template amplification and enrichment were carried out using the standard Ion Torrent protocol using the Ion OneTouch 2 system (Thermo-Fischer). The standard sequencing protocol for Ion Torrent was used and DNA library was loaded onto an Ion316 chip v2 with an ISP loading percentage of 87%.

## **Nanopore sequencing**

Long reads by Nanopore sequencing were obtained using a *MinION* sequencer from Oxford Nanopore Technology (ONT) with the goal to improve the genome assembly of *P. baetica*. The strain was grown in LB medium and grown at 30 °C to an OD<sub>600</sub> of 0.850 and its genomic DNA was isolated using the *Mo Bio DNAeasy UltraClean Microbial* kit. Quality of the gDNA was checked using a NanoDrop and an agarose gel for integrity. The gDNA was sheared mechanically to an average fragment size of 10 kbp using a Covaris *gTube*. The sequencing library was prepared using the 1D ligation protocol from ONT with native barcoding of the sample. The result was sequenced on a R9.4 flowcell for a period of 8 hours.

## **Combined Genome Assembly of Ion PGM and Nanopore Reads**

The Ion PGM reads were checked for quality using *FastQC* and processed with *BBDuk* in order to remove potential adapter contamination, size exclusion (50 > read length < 250), and for trimming (phred score < 28). Basecalling of the Nanopore reads was performed with the software *Albacore* v2.1.3 from ONT and the barcodes were removed using *Porechop*. Read length distribution and quality of the Nanopore reads was assessed using *NanoPlot*.

The genome was first assembled with *SPAdes*, version 3.11.1 using default parameters (Bankevich *et al.* 2012) using Ion PGM reads only, then with *Unicycler*, version v0.4.3 (Wick *et al.* 2017) also using default parameters with both Ion PGM and Nanopore reads. For the latter, the genome was polished with Racon (Vaser *et al.* 2017) and Pilon (Walker *et al.* 2014). The quality of both assemblies was assessed with *QUAST* (Gurevich *et al.* 2013).

## Genomic Island Prediction and Genome Visualisation.

The Ion PGM and the hybrid assemblies were annotated with *Prokka* (Seemann 2014) using a custom protein database built from the *Pseudomonas* genus. The annotations were then searched for prophages using *PHASTER* (Arndt *et al.* 2016), and genomic islands using the *IslandViewer* and antiSMASH (Blin *et al.* 2017; Bertelli *et al.* 2017). For the latter, the reference genome used for the alignment was that of *Pseudomonas koreensis* strain D26 as it was deemed phylogenetically close to *P. baetica*. CCT maps were created using CCT version 1.0 using default settings with the exception of the following commands; -t --cct --custom scale\_blast=F \_cct\_blast\_thickness=60 backboneThickness=10 tickThickness=15 rulerFontSize=100 featureThickness=150 tickLength=30 legend=F details=F. A features file was used for the 'beige blocks' and a 'labels\_to\_show' file was used to label the CCT map (Grant, Arantes and Stothard 2012).

## Phylogenetic analysis

Trees were assembled using Mash v1.1.1 (Ondov *et al.* 2016) fast genome and metagenome distance estimation using the MinHash algorithm, and visualised in FigTree v1.4.3 (Rambaut 2014, also available via GitHub). Reference genomes were obtained from NCBI. Average Nucleotide Identity (ANI) was performed using PyANI (Gupta *et al.* 2016). Data were processed using CLIMB (Connor *et al.* 2016).

## Azocasein Assay

Overnight cultures of *P. aeruginosa* UCBPP-PA14, *P. baetica* a390T, *P. fluorescens* Pf0-1 and *P. putida* KT2440 were centrifuged at 12,000xg for 5 minutes and 100 µl of supernatant was pipetted into centrifuge tubes containing 400 µl of reaction mixture; 200 µl of 2 % azocasein in 0.5 % bicarbonate sodium buffer and 200 µl of 0.5 % sodium bicarbonate buffer. The reaction mixture was then incubated at 30 °C for 10, 70 and 140 minutes.

Following incubation, 0.5 ml of 10 % trichloroacetic acid was added to terminate the reaction. Tubes were then vortexed for 5 minutes followed by centrifuging for 5 minutes at 4 °C and 10,000xg. 500 µl of supernatant was then added to cuvettes containing 500 µl of 0.5

M NaOH and absorbance was measured at 440 nm in an Ultrospec 3100pro spectrophotometer. All samples were run in triplicate.

### Skim Milk Assay

5 µl of overnight cultures of *P. aeruginosa* UCBPP-PA14, *P. baetica* a390T, *P. fluorescens* Pf0-1 and *P. putida* KT2440 were inoculated on to MOPS minimal media agar plates, supplemented with dehydrated skimmed milk and incubated for 72 h at 30 °C. Photographs were taken at 24-hour time intervals. Iron supplement; FeCl<sub>2</sub> was not included in the 10x stock, and 5g/L of agar and skim milk were added in order to observe clear halo formation.

### Oxytetracycline resistance assay

*P. baetica* was tested for antibiotic sensitivity by agar dilution according to the European Committee for Antimicrobial Susceptibility Testing guidelines (European Committee for Antimicrobial Susceptibility Testing (EUCAST) of the European Society of Clinical Microbiology and Infectious Diseases (ESCMID) 2000). Resistance to oxytetracycline was tested due to its frequent use as a growth promoter in aquaculture (Leal *et al.* 2017). Mueller-Hinton agar plates with final concentrations ranging from 0.125 µg mL<sup>-1</sup> to 128 µg mL<sup>-1</sup> were inoculated with *P. baetica* a390T suspensions diluted in 0.85 % NaCl to a turbidity equivalent of a 0.5 McFarland standard. Plates were incubated overnight at 30 °C. *P. aeruginosa* PAO1 was used as a positive control.

## Results

### Sequencing Results and Assembly

The initial Ion Torrent sequencing produced over 3.5 million barcoded reads with a mean read length of 299 bp. SPAdes assembly resulted in final contig number of 338. The Nanopore sequencing run yielded around 22,000 reads, of which 7,771 passed quality control during the basecalling process. The average phred quality score of the reads is 11 and their length distribution was around 9 kbp (see supp. figure 1). With a genome size of about 6.6 Mbp, estimated from an initial Ion PGM assembly, we infer a genome coverage of about 9X for the Nanopore sequencing data (60 Mbp), and 80X with Ion PGM data (540 Mbp after QC).

The hybrid approach that combines both the dataset of short Ion PGM reads with the long Nanopore reads significantly improved the genome assembly over that using Ion PGM reads only giving 16 contigs over 1000 bp (table 1). Quast analysis of the combined assembly gives an N50 value of 973739 which is higher than 96% of all draft genomes in The *Pseudomonas* Genome database.

The Ion PGM-only and the hybrid assembly graphs were also inspected visually using *Bandage* (Wick *et al.* 2015) in order to assess the fragmentation of the genome assemblies and the genomic organisation, including the presence of potential plasmids (see supp. figure 2 and 3). After gene annotation with *Prokka*, plasmid-associated genes like recombinases, transposases, and integrases were searched for and were not present or were not indicative of being plasmid associated. The high continuity assembly provided by the combinatorial approach of Ion Torrent and Nanopore technologies allowed for a comprehensive analysis of the genome.

### **Phylogenetic placement and Average Nucleotide Identity (ANI) of *Pseudomonas baetica* a390T**

We have produced a phylogeny to place the sequenced *P. baetica* a390T genome against reference genomes from 19 species clades, comprising the entire genus *Pseudomonas* (Figure 1B) (Gomila *et al.* 2015). This was done using Mash, a pipeline that converts a genome into a sketch (or groups of k-mers). Sketches are then compared to produce a Jaccard index, which is generated based upon shared k-mers between genomes (Ondov *et al.* 2016). Our phylogeny places *P. baetica* amongst the *P. koreensis* subclade, as previously shown in phylogenies constructed using partial 23S rRNA *P. koreensis* gene sequences. A *P. koreensis* phylogeny constructed in the same manner places *P. baetica* closest to *Pseudomonas* sp. Irchel 3E19, a strain isolated from a pond in Zurich, Switzerland, suggesting that this isolate is in fact a strain of *P. baetica* (Butaitė *et al.* 2017). The next most closely related organism is *P. koreensis* strain CI12, which was recently co-isolated with *Bacillus cereus* from a soybean plant rhizosphere (Bravo, Lozano and Handelsman 2017). ANI was used to compare the whole genome sequence of *P. baetica* against the six available reference genomes for *P. koreensis* as well as strain Irchel 3E19, and confirms that it is most closely related to Irchel 3E19 and CI12, with an ANI of 98.02% and 89.6% similarity, respectively (Figure 1A) (Arahal 2014). We also used JSpeciesWS as an independent ANI method and this analysis agreed with those described above (see supplementary ANI tables) (Richter *et al.* 2016). Finally, BLAST was used to compare the *gyrB* and *rpoD* genes sequences of *P. baetica* a390T and CI12 against available partial *rpoD* and *gyrB* sequences available for *P. baetica* (López *et al.* 2012). This showed a 93.6% similarity in *rpoD* and 95.0% similarity in *gyrB* between CI12 and the whole genome-sequenced *P. baetica* a390T.



## Genomic islands, biosynthetic gene clusters (BGCs) and prophages detected in the *P. baetica* genome

Genomic islands can be readily detected when long contiguous regions are assembled. As such, searching for prophages in the highly fragmented, ion PGM-only assembly did not yield any results, whereas inspection of the hybrid assembly showed 4 intact prophages in the genome of *P. baetica*, as well as 5 incomplete ones (table 2a). Two integrative conjugative elements were also detected. A combination of IslandViewer and antiSMASH were also used to predict genomic islands and the secondary metabolite biosynthetic potential of *P. baetica* (Blin *et al.* 2017; Bertelli *et al.* 2017). These results are summarised on the CG View Comparison Tool map in Figure 2 and a more detailed summary of the antiSMASH results is provided in table 2b (Grant, Arantes and Stothard 2012). Many of the loci referred to throughout this manuscript are labelled around the CCT map in Figure 2 as well as the results of IslandViewer. The CCT BLAST analysis confirms the close phylogenetic relationship between *P. baetica* a390 and *Irchel* 3E19 (Figure 2). Of the nine BGCs predicted by antiSMASH, four are Nrps (Non-ribosomal peptide synthetases) biosynthetic gene clusters. Clusters 3 and 8 had relatively low similarities to the nearest known clusters, however cluster 6 shared a high similarity with an orfamide BGC with 80% of genes in this cluster sharing homology with the putisolvin biosurfactant BGC. Putisolvin has been implicated in motility (Cárcamo-Oyarce *et al.* 2015) and our preliminary experiments have demonstrated that *P. baetica* is capable of both swimming and swarming (see supp. figure 4). Since putisolvin biosynthesis in certain strains of *P. putida* is known to be a quorum sensing (QS)-regulated phenotype, we searched for homologues of known QS systems in the *P. baetica* genome (Dubern *et al.* 2008). No convincing homologues of the LasR or RhIR regulated systems from *P. aeruginosa*, or the PpuL, PpuR, PpuA, RsaL system from *P. putida* were found.

## Bacteriocins, tailocins and modular bacteriocins of *P. baetica*

Pseudomonads are capable of producing a variety of antagonism-mediating peptides and proteins. A subset of these molecules are bacteriocins, and assist in the elimination of phylogenetically-related competitors (Ghequire and de Mot 2014). Historically, these compounds have been mainly studied in *P. aeruginosa*, there termed pyocins (Michel-Briand and Baysse 2002), but they have been equally detected and characterized in a number of other *Pseudomonas* species. Evolutionarily related to bacteriophage tails, tailocins are functional standalone units – rather than defective prophages – lacking an accompanying phage head (Ghequire and de Mot 2015). *Pseudomonas* phage tail-like bacteriocins are typically recruited to *trpE-trpG* and *mutS-cinA* intergenic regions (Ghequire and de Mot 2015). Inspection of the corresponding regions in *P. baetica* reveals the presence of two adjacent R-type gene clusters at the *mutS-cinA* locus. Phylogenetic assessment and gene synteny is indicative of a Rp3-Rp4 organisation, and a similar tailocin configuration was recently described in *Pseudomonas chlororaphis* 30-84 (Dorosky *et al.* 2017). An interesting

observation is that four putative tail fibres are likely present in the Rp4 cassette of *P. baetica*, whereas usually only one or two of these can be detected in tailocins.

Referred to as S-type pyocins in *P. aeruginosa*, modular bacteriocins represent a heterogeneous group of polymorphic toxins (Zhang *et al.* 2012; Jamet and Nassif 2015), comprising a receptor-binding domain, a segment enabling membrane transfer and a toxin domain at the carboxy-terminus (Ghequire and de Mot 2014). To protect from self-inhibition, bacteriocin producer strains co-express an immunity gene, typically encoded adjacent to the bacteriocin killer gene. *Pseudomonas* nuclease bacteriocins (DNase, tRNase, rRNase) are consistently equipped with a central pyocin\_S domain (Pfam06958), suggested to play a pivotal role in translocation of the toxin to the cytoplasm (Ghequire and de Mot 2014; Sharp *et al.* 2017). Searching for this domain revealed an array of putative modular bacteriocins in *P. baetica*, 7 of which host a HNH-DNase and three an rRNase domain. In one bacteriocin a putative (non-HNH-type) DNase homologous to the killer domain found in carocin D – but distinct from the non-HNH-type in pyocin S3 – was identified.

Two other predicted *P. baetica* bacteriocins with homology to a toxin module of a contact-dependent inhibition cassette (type IV) in *Burkholderia pseudomallei* were also detected. No bacteriocins acting at the periplasmic level (colicin M-like, Pfam14859; pesticins, Pfam16754; or pore formers (Colla/ColN), Pfam01024), could be retrieved.

## Lipid A and core oligosaccharide biosynthesis gene clusters

Putative homologues of the confirmed lipid A biosynthesis (Raetz pathway) enzymes in *P. aeruginosa* PAO1 could be identified in the sequenced genome. Comparison to different *Pseudomonas* proteomes revealed sequence identities of the putative proteins ranging from 59 % to 97 %. Those high sequence identities and the hydrocarbon ruler motive in putative LpxA (UDP-N-acetylglucosamine acyltransferase) hint towards a lipid A with a C10 acyl chain at position 3 and 3'

(Smith *et al.* 2015). Besides, C12 secondary acyl chains or C12 primary acyl chains at position 2 and 2' are possible due to sequence homology of putative LpxD and the presents of possible HtrB1 and HtrB2 homologs (Lam *et al.* 2011; Hittle *et al.* 2015). In general, it is likely that *P. baetica* possesses a lipid A structures similar to known structures from *P. aeruginosa* when possible post-synthesis modifications are not taken into account (Knirel *et al.* 2006).

Further genome analysis resulted in the identification of a gene cluster homologous to the core oligosaccharide gene cluster which was described and studied in *P. aeruginosa*. Direct comparison of the cluster revealed that *P. baetica* lacks homologues of the *P. aeruginosa* genes PA4998 (putative kinase) and PA5008 (heptose kinase) which possibly suggests a less phosphorylated core oligosaccharide (Lam *et al.* 2011). Reciprocal BLAST searching of different *Pseudomonas* proteomes resulted in sequence identities between the wide range of 14 % and 98 %. The highest overall sequence similarity was observed to *P. fluorescens*

Pf0-1 core oligosaccharide biosynthesis proteins, ranging from 85 % to 98 % which is consistent with the phylogeny. In general, the cluster structure and presence of putative glycosyltransferases (WapR, WapH, WaaG, WaaC, WaaF) indicates the presence of an inner core oligosaccharide structure similar to *P. aeruginosa* (Knirel *et al.* 2006).

### **Extracellular protease and effector secretion by *P. baetica***

Extracellular protease secretion is an important virulence determinant of *P. aeruginosa*. In order to determine if *P. baetica* has a similar phenotype, azocasein and skimmed milk protease assays were performed with *P. aeruginosa* UCBPP-PA14, *P. baetica*, *P. fluorescens* Pf01 and *P. putida* KT2440. As expected, *P. aeruginosa* UCBPP-PA14 secreted prodigious amounts of protease (see supp. figures 5 & 6). *P. baetica* a390T secreted significantly less protease than *P. aeruginosa* but marginally more than the other strains. This phenotype was most evident on MOPS minimal milk agar after 72 hours (see supp. figure 6b).

A major virulence determinant in Gram-negative pathogens is the ability to inject effector proteins into host cells using a type III secretion system (T3SS) (Hauser 2009). BLAST analysis using the *P. aeruginosa* Psc T3SS proteins revealed a T3SS to be present in the *P. baetica* genome, providing a possible explanation for the virulence of this strain as hypothesised above.

### **Oxytetracycline resistance in *Pseudomonas baetica***

Following overnight incubation, antibiotic plates were examined for bacterial growth and the MIC of oxytetracycline (OT) for *P. aeruginosa* PAO1 was determined as 1  $\mu\text{g mL}^{-1}$ , a quarter that of *P. baetica* a390T. Oxytetracycline is a protein synthesis-inhibiting antibiotic belonging to the tetracycline class, and is one of the most commonly used antibiotics in aquaculture (Leal *et al.* 2017).

When *P. baetica* was cultured on media containing OT it was able to grow in the presence of higher concentrations than *P. aeruginosa* PAO1. The minimum inhibitory concentration of OT required to inhibit the growth of *P. baetica* was determined as 4  $\mu\text{g mL}^{-1}$ . The ability of *P. baetica* to tolerate higher concentrations of OT than a lab strain of *Pseudomonas* may be due to the use of the antibiotic in aquaculture, especially as an additive in fish food. OT resistance (OT<sup>R</sup>) has been observed previously in bacterial species colonising pelletised fish food in Chile, and also in half of commercial fish food samples analysed in a 1995 study of OT<sup>R</sup> in fish farm sediments (Kerry *et al.* 1995; Miranda *et al.* 2003).

## Glycerol-3-Phosphate Utilisation in *P. baetica*

A notable characteristic which differentiates *P. baetica* from closely related *Pseudomonas* species is an inability to grow on glycerol-3-phosphate (G3P) as a sole carbon source (López *et al.* 2012). Several other pseudomonads, such as *P. fluorescens* and *P. aeruginosa* can use G3P as a sole carbon sources. However, *P. baetica* can use glycerol as a sole carbon source. This suggested that *P. baetica* suffers from a specific deficit in G3P uptake, not metabolism. Uptake of G3P is known to occur using the specific MFS transporter GlpT (Fann, Busch and Maloney 2003).

The newly sequenced genome allowed us to examine the *P. baetica* growth deficiency on G3P in greater detail. Remarkably, *P. baetica* appears to harbour a *glpT* homolog, which is quite similar to GlpT from other pseudomonads (see supp. figure 7). The structure of GlpT has been elucidated for *E. coli*, allowing us to understand which residues are important for substrate recognition (Huang *et al.* 2003). All residues which have been identified as important for functionality in previous studies are unchanged in *P. baetica*. Mutation of K80 to alanine killed heterologous G3P-Pi transport of the protein reconstituted into proteoliposomes, and mutation of H165 resulted in a transport rate that was only ~6 % of that catalysed by wild-type protein (Law *et al.* 2008). These data suggest that the presence of a *glpT* homolog does not necessarily equate to G3P uptake, highlighting the necessity of experimental validation. This result may be due to a lack of induction of the *P. baetica glpT* gene upon exposure to G3P.

The broad-spectrum bactericidal antibiotic, fosfomycin, is taken up actively into bacterial cells via GlpT and the glucose-6-phosphate transporter UhpT, and inhibits the initial step in cell wall synthesis (Kahan *et al.* 1974). Inactivation of *glpT* was shown to confer increased fosfomycin resistance in *P. aeruginosa*, with an apparent lack of fitness cost. These authors isolated 10 independent spontaneous mutants, obtained on LB-fosfomycin, which harboured mutations in the *glpT* gene. These mutants included several deletions, frameshifts and three single amino acid changes, Gly<sub>137</sub> to Asp, Thr<sub>336</sub> to Pro and Met<sub>366</sub> to Ile (Castañeda-García *et al.* 2009). Indeed, mutations in the *Staphylococcus aureus glpT* gene were recently shown to be the major determinant of fosfomycin resistance in this organism (Xu *et al.* 2017). The high fosfomycin resistance and low biological fitness cost resulting from the loss of G3P uptake suggests that *P. baetica* may have an evolutionary advantage in a fosfomycin-rich environment. Future work will be necessary to confirm precisely how G3P uptake has been lost in *P. baetica*.

## Orthologs of MexAB-OprM & MexEF-OprN RND efflux systems in *P. baetica*

The MexAB-OprM and MexEF-OprN systems are members of the Resistance-Nodulation-Cell Division (RND) multidrug efflux pump family (Poole 2011). BLASTp was used to

compare the amino acid sequences encoded by the *mexEF-oprN* and *mexAB-oprM* operons in *P. aeruginosa* strain PAO1 with homologues in *P. baetica*.

MexAB-OprM is known to play a key role in efflux-mediated resistance to a wide variety of compounds such as antibiotics and solvents in *P. aeruginosa* (Li, Zhang and Poole 1998; Poole 2011). Similarly, the MexEF-OprN efflux substrates include antibiotics and the quorum sensing precursor 4-hydroxy-2-heptylquinoline (HHQ) (Lamarche and Déziel 2011; Llanes *et al.* 2011). *P. baetica* sequences producing significant alignment to MexAB-OprM and MexEF-OprN were identified with between 69-79 % and 72-88 % identity to *P. aeruginosa* strain PAO1, respectively. Interestingly, the arrangement of the *mexAB-oprM* operon in *P. baetica* bears more similarity to *P. fluorescens*, where the adjacent regulatory gene encodes a member of the TetR regulator family, unlike *P. aeruginosa* which encodes the MarR family regulator MexR in this position. However, it should be noted that *P. baetica* does have a MexR homologue located adjacent to a third efflux operon which has significantly less homology (8-31 %) to *mexAB-oprM*.

### Outer membrane porins in *P. baetica*

In addition to the Mex-associated porins, the *P. baetica* genome was searched for genes coding for other outer membrane porins. There are 26 porin genes in *P. aeruginosa* involved in different important biological functions: outer membrane stability (OprF, OprH) and nutrient uptake (OprB, OprP, and the porins of the OprD family), which have recently been reviewed (Chevalier *et al.* 2017). The *P. baetica* genome revealed only 17 porin genes, but the major porins described in *P. aeruginosa* are conserved. The *oprF* gene coding for the major *Pseudomonas* porin is present in *P. baetica* and the genomic context as well: *estX-menG-cmaX-cfrX-cmpX-sigX-oprF* where EstX is an esterase, MenG a RNase inhibitor, CmaX a CorA-like magnesium transporter, CfrX a putative anti-sigma factor, CmpX a mechano-sensitive channel, and SigX an ECH-sigma factor. OprF has a C-terminal periplasmic domain interacting with peptidoglycan and a barrel N-terminal domain and it is a major contributor to the envelope stability. Other conserved porin genes are *oprG* and *oprH*, which code for small porins with only 8 strands. OprG belongs to the OmpW family and has been proposed to be involved in Fe<sup>2+</sup> diffusion to the periplasm (Catel-Ferreira *et al.* 2016). OprH is induced by low Mg<sup>2+</sup> and contributes to the outer membrane stability via its interaction with LPS (Kucharska *et al.* 2015). As in *P. aeruginosa*, *P. baetica* also has three genes for the glucose porin OprB, including one in the conserved *gltBFGKoprB* operon (Chevalier *et al.* 2017). Surprisingly, the *P. baetica* genome does not contain genes for the OprO and OprP phosphate uptake porins of *P. aeruginosa*. The OprD family comprises 19 members in *P. aeruginosa* sub-divided into two sub-families, the OccD being involved in the uptake of basic amino acids and the OccK for the uptake of negatively charged cyclic molecules (Eren *et al.* 2012). The comparison with *P. aeruginosa* porins via BLASTP is complicated by the fact that many members of the OprD family show significant relatedness. Nevertheless, the survey of the *P. baetica* genome revealed five members of the OccD and six of the OccK family.

## Osmotic stress resistance of *P. baetica*

Since *P. baetica* a390T was originally isolated from a marine aquaculture scenario, we have investigated its genotypic and phenotypic ability to overcome salt stress. Trehalose is a stress-relieving disaccharide which accumulates within the bacterial cell in response to stress. It has been widely studied for its role as a compatible solute that confers protection against osmotic stress. It is thought that trehalose rehydrates the cell following water loss, preventing the loss of cellular functions (De Smet *et al.* 2000). It has also been reported that trehalose is important for pathogenicity of *Pseudomonas* spp. (Freeman, Chen and Beattie 2010; Djonović *et al.* 2013). Bacteria possess three main pathways for trehalose biosynthesis: OtsAB, TreYZ and TreS.

The OtsAB pathway is the most common route of trehalose biosynthesis in bacteria. OtsA catalyses the synthesis of trehalose 6-phosphate from UDP-glucose or ADP-glucose and glucose 6-phosphate. The phosphate group is then removed by OtsB to give trehalose (De Smet *et al.* 2000). The TreYZ and TreS pathways produce trehalose using different substrates. TreY converts reducing maltosyl units of maltooligosaccharides to give  $\alpha$ -1,1 terminal moieties which are then cleaved by TreZ to produce trehalose. The TreS pathway isomerises maltose into trehalose, catalysed by the trehalose synthase TreS, but its utility *in vivo* is context-dependent (Miah *et al.* 2013).

Comparative genomics have shown that in common with other *Pseudomonas* spp., *P. baetica* lacks the genes for the OtsA/OtsB pathway. However, *P. baetica* possesses the enzymes of the TreYZ and TreS pathways.

Although TreS is known to be capable of generating trehalose in some bacteria, it is predicted that the function of TreS in *P. baetica* is to catabolise trehalose. Phenotypic observations suggest that the ability of *P. baetica* to utilise trehalose as a sole carbon source is dependent on other conditions. In a study by López *et al.* it was reported that *P. baetica* was unable to utilise trehalose. Because of its importance to osmotic stress, we repeated the Omnilog analysis for trehalose and observed an intermediate phenotype and further analysis on defined media shows that *P. baetica* is capable of utilising trehalose.

The *treS* gene is fused with the maltokinase-encoding *pep2* gene (Chandra, Chater and Bornemann 2011), consistent with a TreS/Pep2 fusion protein converting trehalose into  $\alpha$ -maltose 1-phosphate (M1P) (Miah *et al.* 2013). M1P is expected to be acted upon by GlgE and GlgB to produce a glycogen-like  $\alpha$ -glucan, given that the *treS/pep2*, *glgE* and *glgB* genes are clustered.

Other *Pseudomonas* spp. such as *P. fluorescens* and *P. syringae* possess an additional homologue of an unfused *treS* gene emphasising the importance of trehalose production in

these organisms. Interestingly this additional homologue is absent in *P. baetica*. *P. fluorescens* and *P. syringae* likely benefit from being able to control the synthesis of trehalose under more varied conditions.

We hypothesised that *P. baetica* was capable of causing disease in marine fish because it had greater salt tolerance than other pseudomonads however, this proved to be incorrect. Our data show that all of the pseudomonads tested here are capable of growth in NaCl concentrations relevant to the marine environment (see supp. figures 8 & 9).

In addition to salt tolerance experiments, *P. baetica* and *P. fluorescens* were also tested for their ability to grow in minimal media with the addition of Instant Ocean which has similar levels of most trace elements, nutrients and alkalinity to sea water samples (Figure 3). *P. baetica* grows well in Instant Ocean while *P. fluorescens* has minimal growth suggesting that *P. baetica* may have some adaptation to the marine environment. Analysis of both genomes using SEED viewer showed that *P. baetica* has a homologue of the *OsmY* gene which has a protein modification function during osmotic stress whereas *P. fluorescens* does not (Métris, George and Ropers 2017).

### Iron uptake genes in *P. baetica*

The *P. baetica* genome was searched for genes coding for siderophore biosynthesis and uptake proteins as well as those involved in  $\text{Fe}^{2+}$  uptake. As expected, genes for the biosynthesis and uptake of the fluorescent siderophore pyoverdine are present (34 in total) and are scattered in four genomic loci as it is usually the case for fluorescent pseudomonads (Ravel and Cornelis 2003). *P. baetica* also produces a second siderophore/metallophore, yersiniabactin, which was described in *Yersinia pestis* and in other Enterobacteriaceae (Chaturvedi *et al.* 2012). Yersiniabactin shares some biosynthesis steps with pyochelin, the second siderophore of *P. aeruginosa* and is able to bind copper with high affinity. It has been also described in some *P. syringae* and *Pseudomonas avellanae* strains (Jones and Wildermuth 2011; Marcelletti and Scortichini 2015). Interestingly, while most pyoverdine genes have their closest homologs in *P. koreensis*, all, but one, yersiniabactin genes of *P. baetica* show a close identity with the yersiniabactin genes from *Pseudomonas azotoformans*, which could be indicative of horizontal gene transfer. The genome contains 27 genes coding for a TonB-dependent receptor, including the *fpvA* gene for the ferripyoverdine uptake, and two *fyuA* genes for the uptake of ferri-yersiniabactin. However, two of these genes are truncated and a probable frameshift is evident for one receptor gene. Noteworthy is the presence of six genes annotated as coding for a TonB protein. As a matter of comparison, only three are present in *P. aeruginosa* (Cornelis, Matthijs and van Oeffelen 2009). Another surprising discovery is the absence of *feoABC* genes for the uptake of  $\text{Fe}^{2+}$  in *P. baetica*. This function is probably secured by the products of the *efeUOAB* operon, first described in *E. coli* O157:H7, which is involved in the uptake of ferrous iron in the periplasm where it gets re-oxidized (Cao *et al.* 2007).

## Discussion

Whole genome sequencing has become ubiquitous in microbiology and access to the technology is now widespread. Here, we show that combining two relatively cheap and user-friendly methods (Ion Torrent Semiconductor and Oxford Nanopore MinION sequencing) enables genome assembly of a relatively large bacterial genome to a finishable state. Our phylogenetic analyses show that *P. baetica* is closely related to *P. koreensis* in the *P. fluorescens* group. However, our data shows that the *P. fluorescens* group taxonomy is in need of review. Given the close relationship of *P. baetica* a390T to *P. irchel* 3e19, we propose that 3e19 is a strain of *P. baetica*. Although both are aquatic isolates, Irchel 3e19 was isolated from fresh water and *P. baetica* a390T is marine (López *et al.* 2012; Butaitė *et al.* 2017). Our data shows that the ability to grow in marine salt concentrations is common in the genus *Pseudomonas*. This, combined with the origin of Irchel 3e19 suggests that *P. baetica* a390T is not a marine restricted organism and is likely to be an opportunist pathogen of marine fish, possibly associated with high fish densities in aquaculture facilities. On the other hand, the presence of a T3SS in both *P. baetica* a390T and Irchel 3e19 (Prof. Rolf Kummerli, personal communication) suggests that these organisms are not merely benign environmental strains. The *P. baetica* genome shows evidence of genome plasticity and contains a variety of genomic islands and prophages as is the case with other pseudomonads including *P. aeruginosa* (Klockgether *et al.* 2011). The diverse iron acquisition and membrane transport systems encoded in the *P. baetica* genome are consistent with an organism that is highly adaptable to a range of environments. We find no genomic evidence of known quorum sensing systems found in other pseudomonads, but our observations that *P. baetica* is capable of protease secretion and swarming suggest that it may engage in social behaviour.

## Acknowledgements

Work in NPT's lab has been funded by BBSRC grant BB/K019600/1 and Chief Scientists Office grant TCS/16/24. AB is supported by a PhD Studentship funded by the Industrial Biotechnology Innovation Centre and the University of Strathclyde. TS was supported by a PhD studentship funded by the Iraqi Government. MG was supported by a postdoctoral fellowship from FWO Vlaanderen (12M4618N). CL was supported by an SB PhD fellowship from FWO Vlaanderen (1S64718N). JB was supported by a grant from Fundação para a Ciência e a Tecnologia (SFRH/BD/104095/2014). NPT would like to personally thank all of the contributors to this paper who have made this novel collaboration an open and pleasurable experience. AM and ED were supported by the Biotechnology and Biological Sciences Research Council-funded South West Biosciences Doctoral Training Partnership (BB/M009122/1).



## References

- Afgan E, Sloggett C, Goonasekera N *et al.* Genomics Virtual Laboratory: A Practical Bioinformatics Workbench for the Cloud. Antoniewski C (ed.). *PLoS ONE* 2015;**10**:e0140829.
- Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. 2014
- Arahal DR. Whole-Genome Analyses: Average Nucleotide Identity. *Methods in Microbiology* 2014;**41**:103–22.
- Arndt D, Grant JR, Marcu A *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016;**44**:W16–21.
- Bankevich A, Nurk S, Antipov D *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;**19**:455–77.
- Bertelli C, Laird MR, Williams KP *et al.* IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res* 2017;**45**:W30–5.
- Blin K, Wolf T, Chevrette MG *et al.* antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* 2017;**45**:W36–W41.
- Bravo JI, Lozano GL, Handelsman J. Draft Genome Sequence of *Flavobacterium johnsoniae* CI04, an Isolate from the Soybean Rhizosphere. *Genome Announc* 2017;**5**:e01535–16.
- Butaitė E, Baumgartner M, Wyder S *et al.* Siderophore cheating and cheating resistance shape competition for iron in soil and freshwater *Pseudomonas* communities. *Nat Commun* 2017;**8**:414.
- Cao J, Woodhall MR, Alvarez J *et al.* EfeUOB (YcdNOB) is a tripartite, acid-induced and CpxAR-regulated, low-pH Fe<sup>2+</sup> transporter that is cryptic in *Escherichia coli* K-12 but functional in *E. coli* O157:H7. *Mol Microbiol* 2007;**65**:857–75.
- Castañeda-García A, Rodríguez-Rojas A, Guelfo JR *et al.* The glycerol-3-phosphate permease GlpT is the only fosfomycin transporter in *Pseudomonas aeruginosa*. *J Bacteriol* 2009;**191**:6968–74.
- Catel-Ferreira M, Marti S, Guillon L *et al.* The outer membrane porin OmpW of *Acinetobacter baumannii* is involved in iron uptake and colistin binding. la Rosa De M (ed.). *FEBS Lett* 2016;**590**:224–31.
- Cárcamo-Oyarce G, Lumjiaktase P, Kümmerli R *et al.* Quorum sensing triggers the stochastic escape of individual cells from *Pseudomonas putida* biofilms. *Nat Commun* 2015;**6**:5945.
- Chandra G, Chater KF, Bornemann S. Unexpected and widespread connections between

- bacterial glycogen and trehalose metabolism. *Microbiology* 2011;**157**:1565–72.
- Chaturvedi KS, Hung CS, Crowley JR *et al.* The siderophore yersiniabactin binds copper to protect pathogens during infection. *Nat Chem Biol* 2012;**8**:731–6.
- Chevalier S, Bouffartigues E, Bodilis J *et al.* Structure, function and regulation of *Pseudomonas aeruginosa* porins. *FEMS Microbiology Reviews* 2017;**41**:698–722.
- Commission E. Aquaculture Facts 2015. 2014:1–1.
- Connor TR, Loman NJ, Thompson S *et al.* CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. *Microb Genom* 2016;**2**:e000086.
- Cornelis P, Matthijs S, van Oeffelen L. Iron uptake regulation in *Pseudomonas aeruginosa*. *Biometals* 2009;**22**:15–22.
- De Smet KA, Weston A, Brown IN *et al.* Three pathways for trehalose biosynthesis in mycobacteria. *Microbiology (Reading, Engl)* 2000;**146** ( Pt 1):199–208.
- Djonović S, Urbach JM, Drenkard E *et al.* Trehalose biosynthesis promotes *Pseudomonas aeruginosa* pathogenicity in plants. Dangl JL (ed.). *PLoS Pathog* 2013;**9**:e1003217.
- Dorosky RJ, Yu JM, Pierson LS III *et al.* *Pseudomonas chlororaphis* produces two distinct R-tailocins that contribute to bacterial competition in biofilms and on roots. Kivisaar M (ed.). *Appl Environ Microbiol* 2017;**83**:AEM.00706–17.
- Dubern J-F, Coppoolse ER, Stiekema WJ *et al.* Genetic and functional characterization of the gene cluster directing the biosynthesis of putisolvin I and II in *Pseudomonas putida* strain PCL1445. *Microbiology (Reading, Engl)* 2008;**154**:2070–83.
- Eren E, Vijayaraghavan J, Liu J *et al.* Substrate specificity within a family of outer membrane carboxylate channels. Dutzler R (ed.). *PLoS Biol* 2012;**10**:e1001242.
- European Committee for Antimicrobial Susceptibility Testing (EUCAST) of the European Society of Clinical Microbiology and Infectious Diseases (ESCMID). EUCAST Definitive Document E.DEF 3.1, June 2000: Determination of minimum inhibitory concentrations (MICs) of antibacterial agents by agar dilution. *Clin Microbiol Infect* 2000;**6**:509–15.
- Fann M-C, Busch A, Maloney PC. Functional characterization of cysteine residues in GlpT, the glycerol 3-phosphate transporter of *Escherichia coli*. *J Bacteriol* 2003;**185**:3863–70.
- Freeman BC, Chen C, Beattie GA. Identification of the trehalose biosynthetic loci of *Pseudomonas syringae* and their contribution to fitness in the phyllosphere. *Environmental Microbiology* 2010;**12**:1486–97.
- Ghequire MGK, de Mot R. Ribosomally encoded antibacterial proteins and peptides from *Pseudomonas*. *FEMS Microbiology Reviews* 2014;**38**:523–68.
- Ghequire MGK, de Mot R. The Tailocin Tale: Peeling off Phage Tails. *Trends Microbiol*

2015;**23**:587–90.

Gomila M, Peña A, Mulet M *et al.* Phylogenomics and systematics in *Pseudomonas*. *Front Microbiol* 2015;**6**:214.

Grant JR, Arantes AS, Stothard P. Comparing thousands of circular genomes using the CGView Comparison Tool. *BMC Genomics* 2012;**13**:202.

Gupta V, Haider S, Sood U *et al.* Comparative genomic analysis of novel *Acinetobacter* symbionts: A combined systems biology and genomics approach. *Sci Rep* 2016;**6**:29043.

Gurevich A, Saveliev V, Vyahhi N *et al.* QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;**29**:1072–5.

Hauser AR. The type III secretion system of *Pseudomonas aeruginosa*: infection by injection. *Nat Rev Micro* 2009;**7**:654–65.

Hittle LE, Powell DA, Jones JW *et al.* Site-specific activity of the acyltransferases HtrB1 and HtrB2 in *Pseudomonas aeruginosa* lipid A biosynthesis. Carbonetti N (ed.). *Pathog Dis* 2015;**73**:ftv053.

Huang Y, Lemieux MJ, Song J *et al.* Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. *Science* 2003;**301**:616–20.

Jamet A, Nassif X. New players in the toxin field: polymorphic toxin systems in bacteria. *MBio* 2015;**6**:e00285–15.

Joh S-J, Ahn E-H, Lee H-J *et al.* Bacterial pathogens and flora isolated from farm-cultured eels (*Anguilla japonica*) and their environmental waters in Korean eel farms. *Vet Microbiol* 2013;**163**:190–5.

Jones AM, Wildermuth MC. The phytopathogen *Pseudomonas syringae* pv. tomato DC3000 has three high-affinity iron-scavenging systems functional under iron limitation conditions but dispensable for pathogenesis. *J Bacteriol* 2011;**193**:2767–75.

Kahan FM, Kahan JS, Cassidy PJ *et al.* The mechanism of action of fosfomycin (phosphonomycin). *Ann N Y Acad Sci* 1974;**235**:364–86.

Kerry J, Hiney M, Coyne R *et al.* Fish feed as a source of oxytetracycline-resistant bacteria in the sediments under fish farms. *Aquaculture* 1995;**131**:101–13.

Keshavarz-Tohid V, Taheri P, Muller D *et al.* Phylogenetic diversity and antagonistic traits of root and rhizosphere pseudomonads of bean from Iran for controlling *Rhizoctonia solani*. *Res Microbiol* 2017;**168**:760–72.

Klockgether J, Cramer N, Wiehlmann L *et al.* *Pseudomonas aeruginosa* Genomic Structure and Diversity. *Front Microbiol* 2011;**2**:150.

Knirel YA, Bystrova OV, Kocharova NA *et al.* Review: Conserved and variable structural

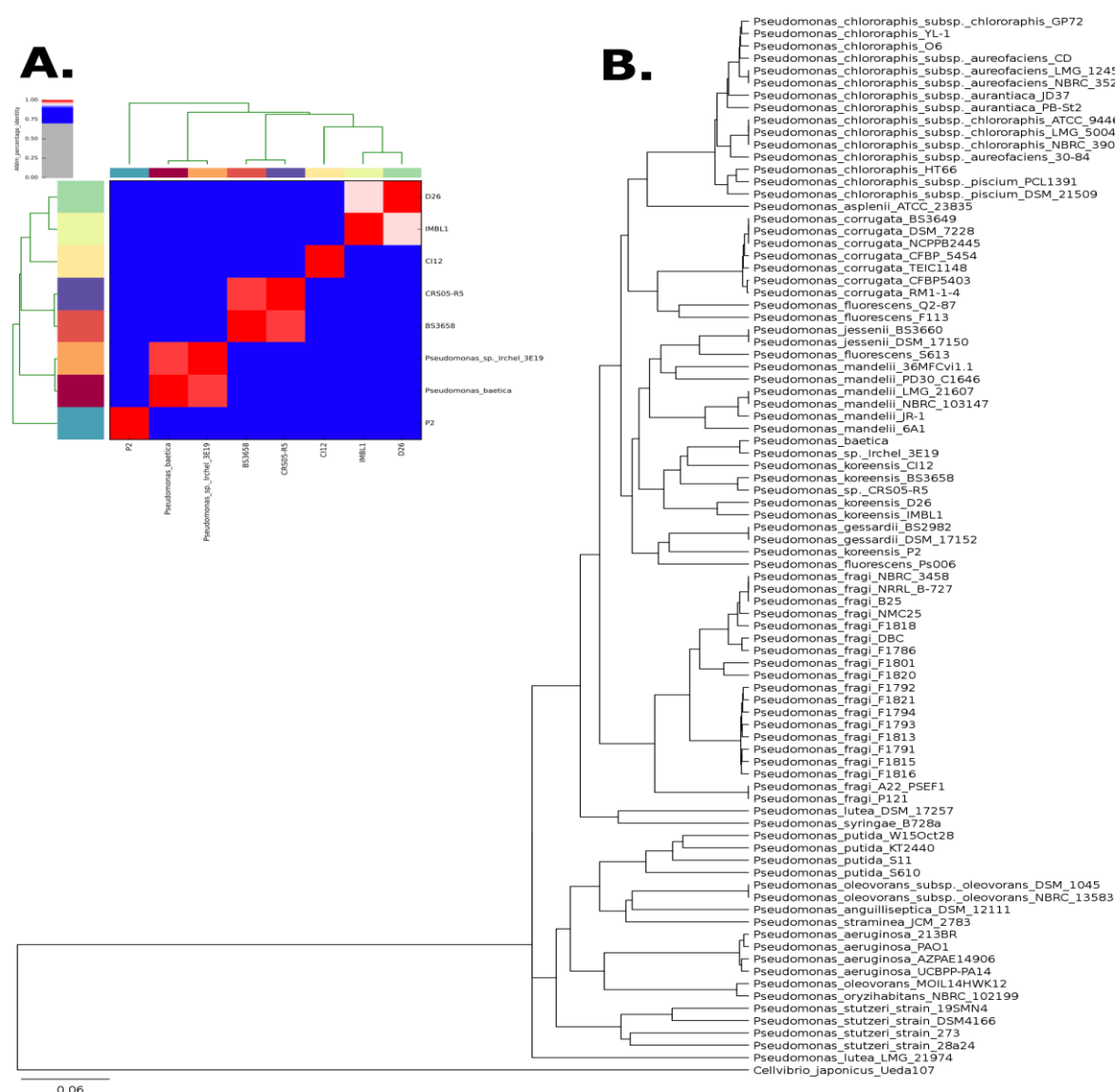
- features in the lipopolysaccharide of *Pseudomonas aeruginosa*. *Journal of Endotoxin Research* 2006;**12**:324–36.
- Kucharska I, Seelheim P, Edrington T *et al*. OprG Harnesses the Dynamics of its Extracellular Loops to Transport Small Amino Acids across the Outer Membrane of *Pseudomonas aeruginosa*. *Structure* 2015;**23**:2234–45.
- Lam JS, Taylor VL, Islam ST *et al*. Genetic and Functional Diversity of *Pseudomonas aeruginosa* Lipopolysaccharide. *Front Microbiol* 2011;**2**:118.
- Lamarche MG, Déziel E. MexEF-OprN efflux pump exports the *Pseudomonas* quinolone signal (PQS) precursor HHQ (4-hydroxy-2-heptylquinoline). Otto M (ed.). *PLoS ONE* 2011;**6**:e24310.
- Law CJ, Almquist J, Bernstein A *et al*. Salt-bridge dynamics control substrate-induced conformational change in the membrane transporter GlpT. *J Mol Biol* 2008;**378**:828–39.
- Leal JF, Henriques IS, Correia A *et al*. Antibacterial activity of oxytetracycline photoproducts in marine aquaculture's water. *Environ Pollut* 2017;**220**:644–9.
- Li XZ, Zhang L, Poole K. Role of the multidrug efflux systems of *Pseudomonas aeruginosa* in organic solvent tolerance. *J Bacteriol* 1998;**180**:2987–91.
- Llanes C, Köhler T, Patry I *et al*. Role of the MexEF-OprN efflux system in low-level resistance of *Pseudomonas aeruginosa* to ciprofloxacin. *Antimicrob Agents Chemother* 2011;**55**:5676–84.
- López JR, Diéguez AL, Doce A *et al*. *Pseudomonas baetica* sp. nov., a fish pathogen isolated from wedge sole, *Dicologlossa cuneata* (Moreau). *Int J Syst Evol Microbiol* 2012;**62**:874–82.
- López JR, Lorenzo L, Marcelino-Pozuelo C *et al*. *Pseudomonas baetica*: pathogenicity for marine fish and development of protocols for rapid diagnosis. *FEMS Microbiol Lett* 2017;**364**:fnw286.
- Marcelletti S, Scortichini M. Comparative Genomic Analyses of Multiple *Pseudomonas* Strains Infecting *Corylus avellana* Trees Reveal the Occurrence of Two Genetic Clusters with Both Common and Distinctive Virulence and Fitness Traits. Semsey S (ed.). *PLoS ONE* 2015;**10**:e0131112.
- Métris A, George SM, Ropers D. Piecewise linear approximations to model the dynamics of adaptation to osmotic stress by food-borne pathogens. *Int J Food Microbiol* 2017;**240**:63–74.
- Miah F, Koliwer-Brandl H, Rejzek M *et al*. Flux through trehalose synthase flows from trehalose to the alpha anomer of maltose in mycobacteria. *Chem Biol* 2013;**20**:487–93.
- Michel-Briand Y, Baysse C. The pyocins of *Pseudomonas aeruginosa*. *Biochimie* 2002;**84**:499–510.

- Miranda CD, Kehrenberg C, Ulep C *et al.* Diversity of tetracycline resistance genes in bacteria from Chilean salmon farms. *Antimicrob Agents Chemother* 2003;**47**:883–8.
- Ondov BD, Treangen TJ, Melsted P *et al.* Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* 2016;**17**:132.
- Pascual J, García-López M, Bills GF *et al.* *Pseudomonas granadensis* sp. nov., a new bacterial species isolated from the Tejeda, Almijara and Alhama Natural Park, Granada, Spain. *Int J Syst Evol Microbiol* 2015;**65**:625–32.
- Perkel JM. How scientists use Slack. *Nature* 2016;**541**:123–4.
- Poole K. *Pseudomonas aeruginosa*: resistance to the max. *Front Microbiol* 2011;**2**:65.
- Rambaut A. FigTree v1. 4. 2012 <http://tree.bio.ed.ac.uk/software/figtree/> (7 December 2017, date last accessed).
- Ramírez-Bahena M-H, Cuesta MJ, Flores-Félix JD *et al.* *Pseudomonas helmanticensis* sp. nov., isolated from forest soil. *Int J Syst Evol Microbiol* 2014;**64**:2338–45.
- Ravel J, Cornelis P. Genomics of pyoverdine-mediated iron uptake in pseudomonads. *Trends Microbiol* 2003;**11**:195–200.
- Richter M, Rosselló-Móra R, Oliver Glöckner F *et al.* JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 2016;**32**:929–31.
- Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;**30**:2068–9.
- Sharp C, Bray J, Housden NG *et al.* Diversity and distribution of nuclease bacteriocins in bacterial genomes revealed using Hidden Markov Models. Orengo CA (ed.). *PLoS Comput Biol* 2017;**13**:e1005652.
- Smith EW, Zhang X, Behzadi C *et al.* Structures of *Pseudomonas aeruginosa* LpxA Reveal the Basis for Its Substrate Selectivity. *ACS Publications* 2015;**54**:5937–48.
- Vaser R, Sović I, Nagarajan N *et al.* Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* 2017;**27**:737–46.
- Walker BJ, Abeel T, Shea T *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. Wang J (ed.). *PLoS ONE* 2014;**9**:e112963.
- Wick RR, Judd LM, Gorrie CL *et al.* Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. Phillippy AM (ed.). *PLoS Comput Biol* 2017;**13**:e1005595.
- Wick RR, Schultz MB, Zobel J *et al.* Bandage: interactive visualization of de novo genome

assemblies. *Bioinformatics* 2015;**31**:3350–2.

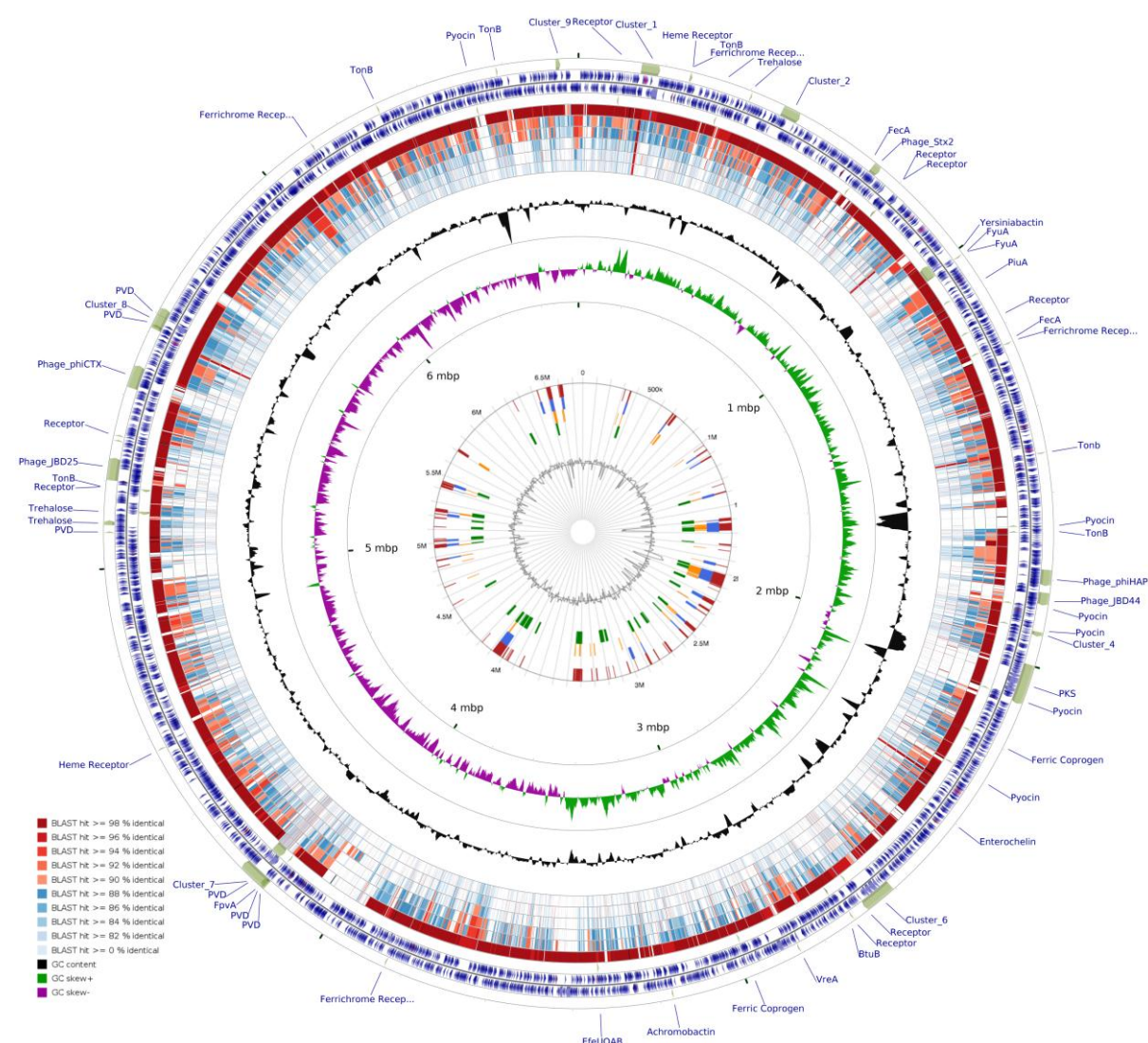
Xu S, Fu Z, Zhou Y *et al.* Mutations of the Transporter Proteins GlpT and UhpT Confer Fosfomycin Resistance in *Staphylococcus aureus*. *Front Microbiol* 2017;**8**:914.

Zhang D, de Souza RF, Anantharaman V *et al.* Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol Direct* 2012;**7**:18.



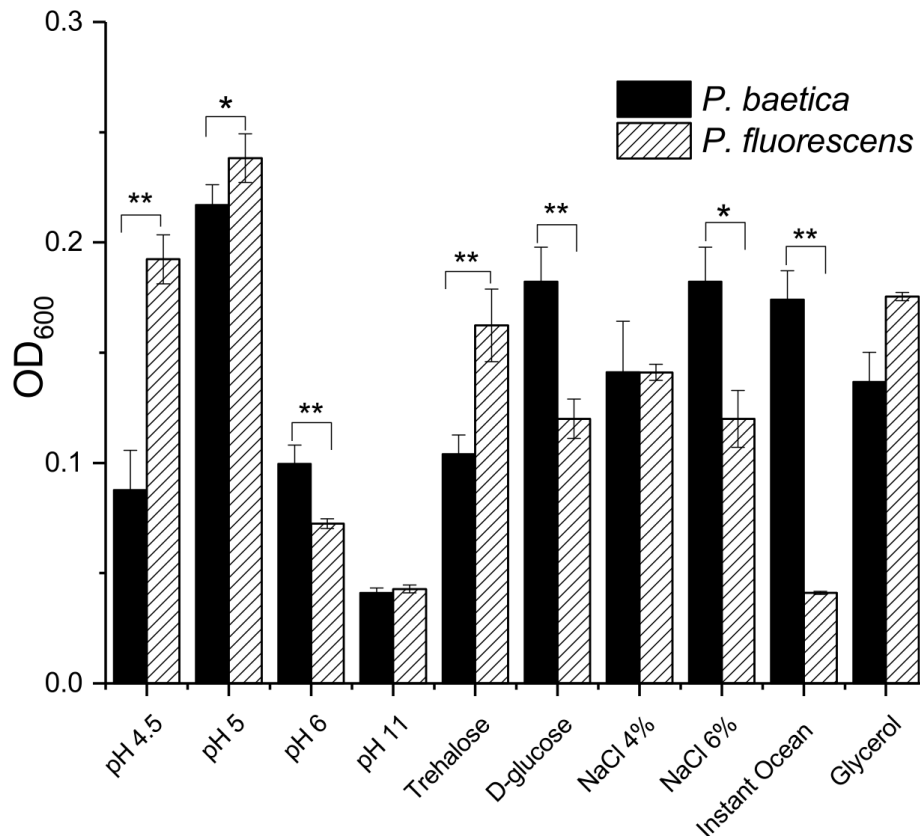
**Figure 1. Placement of novel *baetica* genome in the genus *Pseudomonas* based upon Average Nucleotide Identity, and shared k-mers A) Average Nucleotide Identity (ANI) heatmap generated the Python3 module pyani. The sequenced *baetica* genome, *Pseudomonas\_sp\_Irchel\_3E19* and reference genomes from the closely related *koreensis* subclade were subject to ANI analysis. B) A Mash-based tree generated from reference genomes ( $n = 86$ ) from 19 species clades, comprising the entire genus *Pseudomonas*. This tree**

was generated based upon the Jaccard index, calculated from shared k-mers. *Cellvibrio japonicum* Ueda107 was used as outgroup.



**Figure 2. Overview of the genomic architecture of *Pseudomonas baetica* a390T.** CGView Comparison Tool was used to plot the percentage sequence identity of closely related strains. The outer labels and beige blocks refer to loci that are mentioned throughout the paper. The *P. baetica* a390T genome is split into the forward and reverse strands and the coding sequences are represented by blue arrows. The next 6 rings represent the percentage BLAST identity of the genomes of the following closely related strains; from the outside ring to inside the order is *Pseudomonas sp* Irchel 3E19, *Pseudomonas koreensis* CI12, *Pseudomonas koreensis* D26, *Pseudomonas fluorescens* SBW25, *Pseudomonas putida* KT2440, *Pseudomonas aeruginosa* UCBPP-PA14. The percentage identity is indicated by the colour of the 6 BLAST rings as indicated by the key on the bottom left hand side of the figure. Moving inwards, the next two rings indicate the changes in %GC content and the GC skew respectively. Finally, the central diagram indicates the genomic islands predicted by the IslandViewer package where Red = predicted by multiple methods, blue = IslandPath-DIMOB, orange = SIGI-HMM and green = IslandPick.





**Figure 3. Stress tolerance and growth characteristics of *P. baetica* a390T.** Graph of end point growth measured as OD<sub>600</sub> after 24 hours incubation at 30°C in *Pseudomonas* minimal media (L-glutamine 5%, K<sub>2</sub>HPO<sub>4</sub> 1.5%, MgSO<sub>4</sub> 0.2%, 20mM glycerol or other carbon source as indicated) with pH or salt concentration adjusted as indicated. Data represented mean ± SD. \**p* < 0.05, \*\**p* < 0.01 (determined by two-way T-test).

Table 1: assessment of genome assembly quality for both approaches reveals a vast improvement in the assembly despite a low nanopore sequencing coverage (9X).

Metric	Ion PGM only	Hybrid
Assembly length	6,602,908	6,773,804
#contigs (>1000bp)	338	16
N50	36,090	973,739
N75	19,947	745,006
Longest contig	122,046	1,654,292



Table 2a. List of prophage regions detected in the hybrid assembly.

contig#	Length	Completeness	Score	# Proteins	Region Position	Most Common Phage	GC %
1	6.7 Kbp	questionable	70	6	681684-688458	PHAGE_Stx2_c_1717_NC_011357(3)	53.77%
2	6.5 Kbp	incomplete	40	10	512098-518615	PHAGE_Enterо_phi92_NC_023693(2)	60.11%
2	10.2 Kbp	incomplete	30	8	742363-752628	PHAGE_Clostr_phiCT453A_NC_028991(3)	56.63%
2	5.5 Kbp	incomplete	60	6	925750-931322	PHAGE_Stx2_c_1717_NC_011357(3)	54.15%
3	40.7 Kbp	intact	150	48	383269-424017	PHAGE_Vibrio_VP882_NC_009016(12)	58.74%
3	57.2 Kbp	intact	150	49	430072-487323	PHAGE_Vibrio_vB_VpaM_MAR_NC_019722(8)	58.69%
4	35.1 Kbp	intact	150	42	357957-393153	PHAGE_Pseudo_phiCTX_NC_003278(18)	57.32%
4	50.1 Kbp	intact	150	70	564492-614686	PHAGE_Enterо_Arya_NC_031048(10)	56.98%
5	21.3 Kbp	incomplete	40	11	26364-47679	PHAGE_Pseudo_vB_PsyM_KIL1_NC_030934(6)	49.79%
5	8.7 Kbp	incomplete	50	13	540790-549505	PHAGE_Bacill_BMBtpLA_NC_028748(1)	57.37%

Table 2b. Secondary metabolite biosynthetic clusters predicted by antiSMASH

Cluster	Type	Cluster Size (bp)	Most Similar Known Biosynthetic Gene Cluster	Percentage Similarity
1	Unknown	43395	Mangotoxin	71
2	Arylpolyene	43613	APE_Vf	40
3	Nrps	63804	Nataxazole	11
4	Bacteriocin/Ripp	10887	-	-
5	Transatpks-Nrps	94211	Sorangicin	13
6	Nrps	77485	Orfamide / Putisolvin	70
7	Terpene-Nrps	73119	Pyoverdine	16
8	Nrps	52998	Pyoverdine	22
9	Bacteriocin/Ripp	10845	-	-